



Title	An EMA study of the articulatory-acoustic relationship of Cantonese corner vowels
Author(s)	Lam, Kin-man; 林健文
Citation	
Issued Date	2009
URL	http://hdl.handle.net/10722/173693
Rights	Creative Commons: Attribution 3.0 Hong Kong License

An EMA study of the articulatory-acoustic relationship of Cantonese corner vowels

Lam Kin Man

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science
(Speech and Hearing Sciences), The University of Hong Kong, June 30, 2009.

Abstract

The study examined the articulatory-acoustic relationship of corner vowel productions.

Electromagnetic Articulography (EMA) was used to measure the articulatory parameters as expressed in terms of location as a function of time. The kinematic measures were correlated with acoustic parameters including the first and second formant frequencies. Results indicated correlation between articulatory and acoustic Euclidean distances was weak for males, but moderately strong for females. Comparing the acoustic and physiological vowel spaces of various Cantonese vowels, the distance between /u/ and /a/ was relatively shorter in acoustic than in articulation. The implications of gender difference on acoustic-articulatory relationship were also discussed.

Introduction

Vowel identity can be acoustically described in terms of formant frequencies, based on which different vowels can be distinguished from each other. Formants are the energy peaks found in the output speech resulted from oral, nasal, and pharyngeal resonances in the vocal tract during speech production. According to the source-filter theory, a change in the vocal tract configuration leads to a change in the resonance characteristics of the vocal tract, and as a result, formants change. Formant change therefore is a strong indication of any articulatory movement inside the vocal tract (Gloria, Harris & Raphael, 1994). Studies have reported that the first formant (F1) generally decreases with jaw and tongue height, and the second formant frequency (F2) increases with tongue frontness (e.g., Johnson, 2003).

A number of articulatory studies have been carried out with the assumption of the source-filter theory. Kent, Kent, Rosenbek, Weismer, Martin, Sufit and Brooks (1992) investigated dysarthric speech and observed a reduction of F1 and F2 spectral slope in dysarthric speech. Based on this they inferred that there was a reduced range of articulation in dysarthric patients. This conclusion was drawn assuming a simple and linear relationship between acoustics and articulation, which is in line with the assumption of the source-filter theory. However, several studies have countered such simple and linear assumption. Hashi, Kent, Westbury and Lindstrom (1996) reported that acoustics and articulation are in fact not linearly related. This implies that the relationship of F1-tongue height and F2-tongue

advancement may not be plausible. Some of their subjects even exhibited an unusual acoustic-articulatory relationship: a decreased F2 was associated with forward movement of the tongue. This finding clearly contradicts the traditionally claimed positive correlation between F2 and tongue frontness.

The unusual correlation of F2 change and tongue movement was also reported in recent studies. Ling (2007) reported that Suzhou fricative vowel has a lower F2 but more anterior tongue constriction when compared to its counterpart vowel without frication. Lee (2005) also reported a lower F2 value associated with apical anterior vowel when compared with apical posterior vowel. All of these studies argued that formant values may not have a linear relationship with tongue position (vocal tract configuration). However, all of these studies failed to quantify for or formulate a new acoustic-articulatory relationship.

In an attempt to validate the linear relationship between acoustics and articulation, Eir Cortes and Lindblom (2008) quantified tongue height and correlated tongue height measurement with F1 values. They found poor correlation between F1 and jaw height, with correlation coefficients of 0.17 for vowels in /b/-context, and 0.31 for vowels in /g/-context. However, the study did not provide information on F2 values. The close tongue-jaw interaction in vowel production was also not considered, when the jaw height was measured in isolation of the tongue movement. The study failed to provide a full account of the acoustic-articulatory correlation.

Quite consistently, results from previous studies do not seem to support the simple linear relationship between acoustics (represented by formants) and articulation (represented by positioning of articulators). Researchers begin to suggest a more complex and non-linear relationship between acoustics and articulation. However, discussion so far has been vague and lacks concrete quantitative evidence. In addition, studies supporting the notion of non-linear acoustic-articulatory relationship were mostly limited to comparison of neighboring speech sounds, or tracking a single word with time (Eir Cortes & Lindblom, 2008; Hashi, Kent, Westbury & Lindstrom, 1996; Lee, 2005; Ling, 2007). The acoustic and articulatory changes observed were generally small in scale. The acoustic-articulatory relationship would easily be altered by other minute oromotor changes which may not be detected and considered in these studies. Vowels with greater contrast in tongue place and height are expected to illustrate the acoustic-articulatory relationship more clearly. In the present study, the relationship between formant frequencies and articulatory movements was quantified among the three corner vowels in Cantonese /i, a, u/. In order to obtain more accurate first-hand articulatory data, electromagnetic articulography (EMA) was used.

Based on the theory of electromagnetism, EMA makes use of alternating magnetic fields to induce electric current in sensor coils attached to various speech articulators. By measuring the strength of the induced electric current, the spatial configuration of articulators to which the sensors are attached can be determined (Perkell, Cohen, Svirsky, Matthies,

Garabieta & Jackson, 1992). Previously, researchers have attempted to use x-ray technique and ultrasonography to investigate articulatory movement. Compared to x-ray techniques, EMA is non-invasive and therefore it allows researchers to include more participants and more speech samples from each participant, by doing so researchers can balance the significant intra-subject and inter-subject variability in speech movements (Goozée, Lapointe & Murdoch, 2003). Ultrasonic imaging is also not invasive; however, EMA possesses a higher frame rate (sampling rate = 200 Hz) which is capable of accurately detecting the fast and dynamic real-time speech movements (Kaburagi & Honda, 2002).

Method

Participants

A total of 10 participants (five males and five females) were recruited. Since Simpson (2002) reported significant gender effect in articulatory-acoustic relations with female speech acoustically more sensitive to articulation, the participants were divided into two separate gender groups. The speakers were with ages ranging from 20 to 22 years. All of them were native Cantonese speakers with no perceptual speech errors.

Speech Stimuli

To obtain productions of the three corner vowels (/i/, /a/ and /u/), each subject produced each target word in a carrier phrase ten times at a randomized order. There were three target words: /pin1/ (邊 / border), /pan1/ (班 / class) and /pun1/ (搬 / carry). Their

common phonemic context balanced the influence of adjacent consonants on the vowel productions. All three target words were in high-level tone, so as to minimize the possible effect of tonal variation with vowel production (Hoole & Hu, 2004). To simulate the articulations in conversation, the speakers produced the target words in a carrier phrase: /kən1 tsy6 tuk9 __ pei2 nei5 tʰɛŋ1/ (跟住讀____俾你聽) meaning “Then, I will read ____ to you”). Each target word was repeated ten times to balance the time-to-time variability in articulations. To reduce the influence of fatigue among the vowel productions and the possible order effect, all speech materials were presented in a randomized order.

Equipments and Data Collection

Articulatory data was acquired by using a 3-D electromagnetic articulography (EMA) (AG500, Carstens Medizinelektronik GmbH). During data recording, the speaker was seated in a frame which generated an alternating magnetic field. Electric current was induced in the recording sensors that were attached to articulators upon movement. Based on the magnitude of induced electric current, the change of sensor position was determined. In this study, three sensors were attached on the tongue at 1 cm, 2cm, and 3 cm from the tongue tip. Additional reference sensors were attached to the nose bridge, the gum above the upper central incisors, and behind the ear, so that the software Norm-Pos was able to detect and cancel out the effect of head movement. To familiarize them with the recording format, the participants were provided a practice period with the speech materials and casual conversation with the

experimenter after the sensors had been attached to the articulators. Only those participants who were able to produce speech with ease were allowed to continue with the experiment.

Acoustic data was collected by using a unidirectional dynamic vocal microphone recorded simultaneously with articulation data. The acoustic signals were digitized at a sampling rate of 10 kHz and used for later acoustic analysis.

Data Analysis

With regard to articulatory data, for each vowel sample, three data points were obtained for each plane, originating from the three lingual sensors. To reduce the amount of data, the centroid of these three data points was calculated and used to represent the tongue location during vowel production for each plane. As a result, only one datum for each plane was derived from the three lingual sensors and it was used to represent the tongue location. Since only tongue height and tongue advancement were of interest in the present study, only the vertical and horizontal planes were considered.

The first and second formant frequencies (F1 & F2) were obtained from the acoustic signals by using Praat, which is a signal analysis software. The waveform and the corresponding wideband spectrogram of each acoustic signal were displayed on the screen, and F1 and F2 values were calculated from the mid-point of each vowel by using Linear Prediction Coding (LPC) of Praat.

For both articulatory and acoustic data, measurements were separated into discrete

horizontal and vertical dimensions. This was based on the assumption that articulation only interacts with acoustics in the same plane, which may not be true. Hence, a quantitative measure known as Euclidean Distance (ED) was used to calculate the acoustic and articulatory “distances” between any two vowels. The calculation of ED was based on the Pythagorean Theorem in which horizontal and vertical measurements resembled the width and height of a right-angled triangle. ED was calculated using both acoustic and articulatory data, yielding acoustic ED and articulatory ED.

Upon completion of ED calculation, acoustic ED and articulatory ED values were correlated. Pearson product-moment correlation coefficients were calculated to assess the associative relationship between acoustics and articulation. In addition to ED calculation, vowel spaces were also constructed using both acoustic and articulatory data. The articulatory vowel space was constructed by plotting the tongue frontness on the horizontal axis, and the tongue height on the vertical axis. The acoustic vowel space was constructed by plotting the F2 on the horizontal axis, and negative F1 on the vertical axis. Negative F1 was used due to the inversely proportional relationship between F1 and tongue height. This allowed the comparison of the acoustic and articulatory vowel spaces.

Results

Euclidean Distance Comparison

Table 1 depicts the acoustic and articulatory Euclidean Distances (EDs) of the three

corner vowels. Pearson product-moment correlation between acoustic and articulatory EDs was weak for males ($r = -0.255$, $p < 0.001$), but moderately strong for females ($r = -0.484$, $p < 0.001$).

Table 1. Mean (M), standard deviation (SD) and range values of Euclidean distances (ED) in articulation and acoustics for the three corner vowels produced by male and female native Cantonese speakers.

	Males (n=150)			Females (n=150)		
	M	SD	Range	M	SD	Range
Articulatory ED (mm)	73.49	18.40	4.71–135.07	57.93	6.90	46.18–104.39
Acoustic ED (Hz)	1815.45	532.64	930.45–2684.54	1918.34	605.20	972.12–2855.81

Vowel Space Comparison

As shown in Figures 1 and 2, the acoustic and articulatory vowel spaces were superimposed on each other for easy comparison. The two vowel spaces were scaled in such a way that the distance between the vowel /i/ and the centroid of vowel space in both vowel spaces were similar. It is noted that the articulatory vowel space exhibited shorter /i/-/a/ distance, and longer /u/-/a/ distance when compared with acoustic vowel space, regardless of gender.

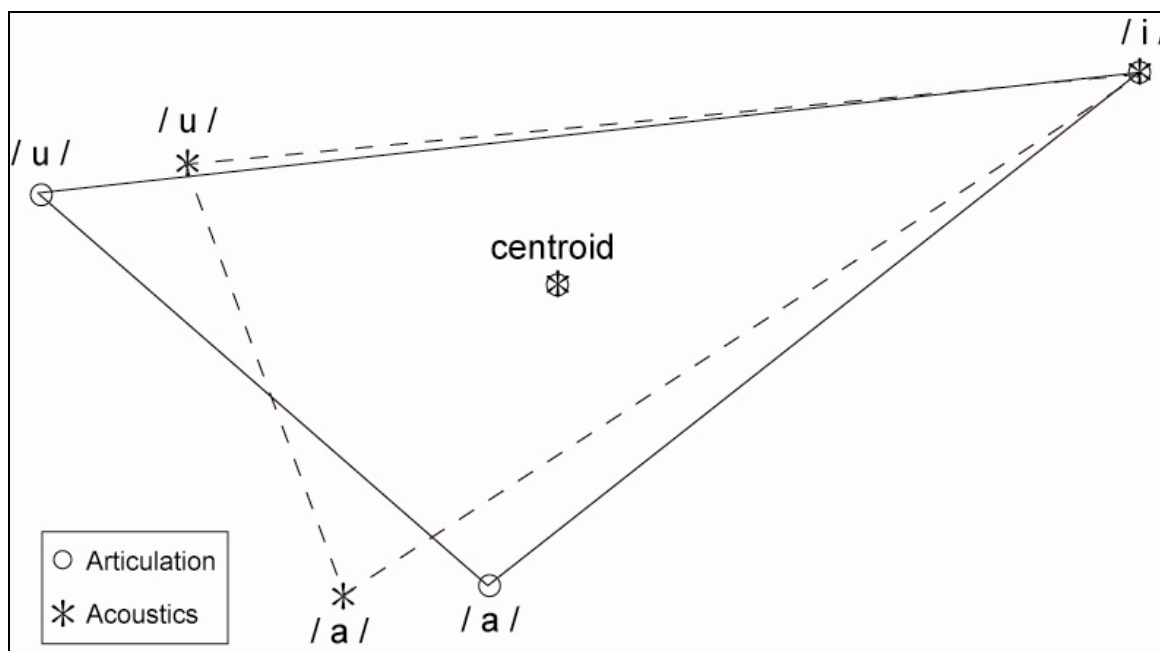


Figure 1. Acoustic and articulatory vowel spaces for males.

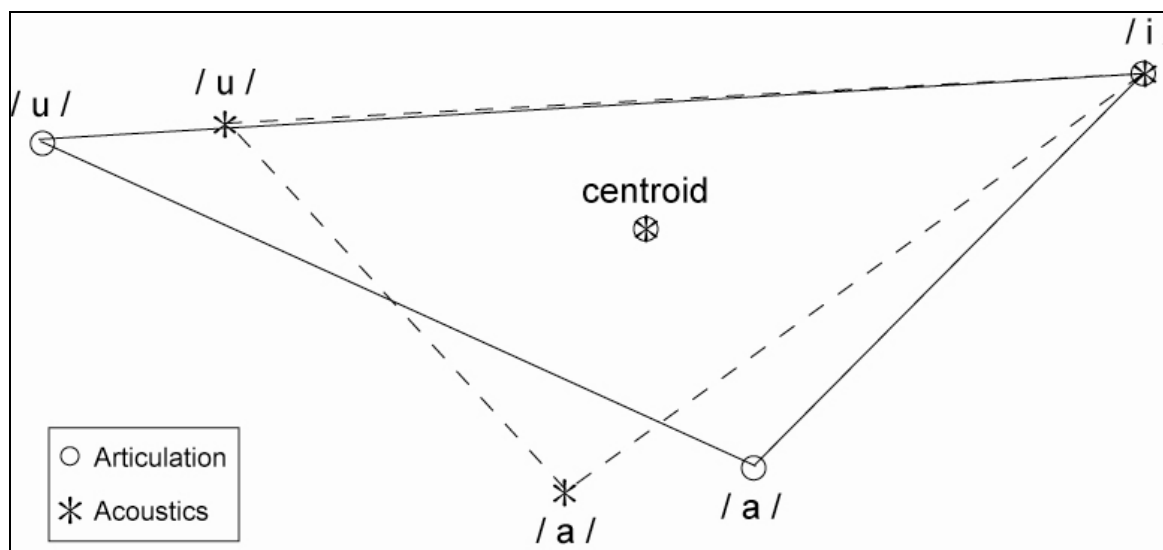


Figure 2. Acoustic and articulatory vowel spaces for females.

Gender Comparison

Articulatory and acoustic vowel spaces for males and females are shown in Figures 3

and 4, respectively. From Figure 3, it should be noted that both male and female articulatory vowel spaces were similar and comparable in height, except that female articulatory vowel space appeared to be slightly shorter than male articulatory vowel space. Figure 4 shows that female's acoustic vowel space was of a larger scale in both F1 and F2 dimensions than male's.

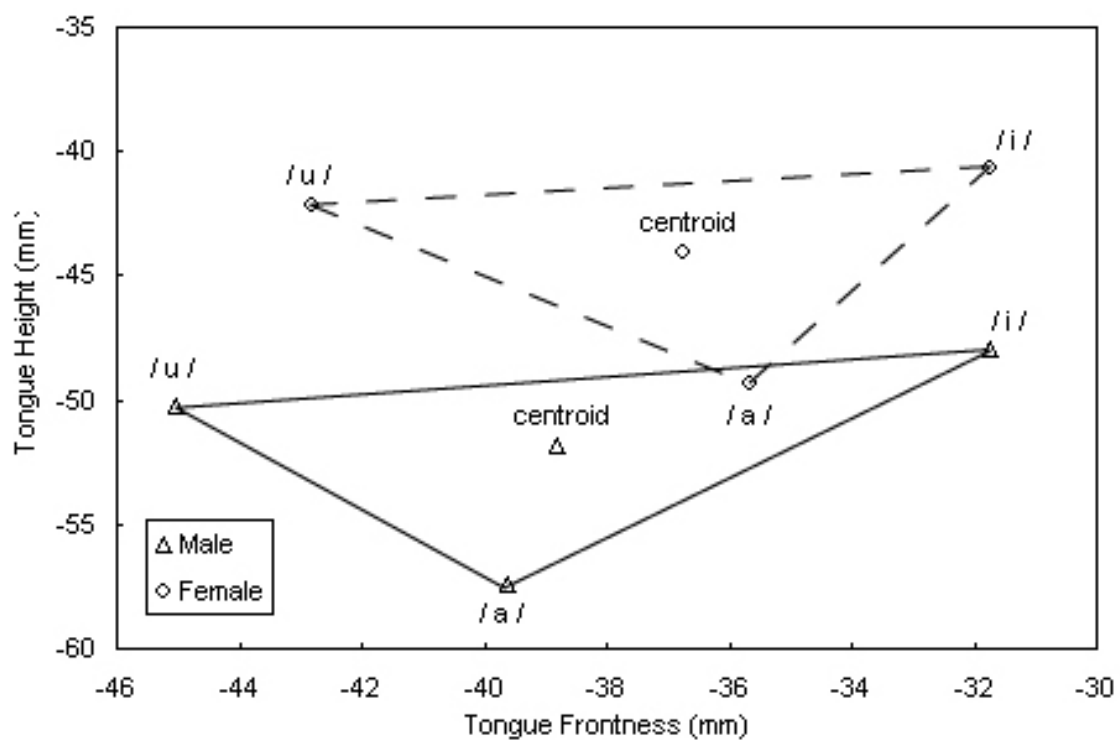


Figure 3. Articulatory vowel spaces for males and females.

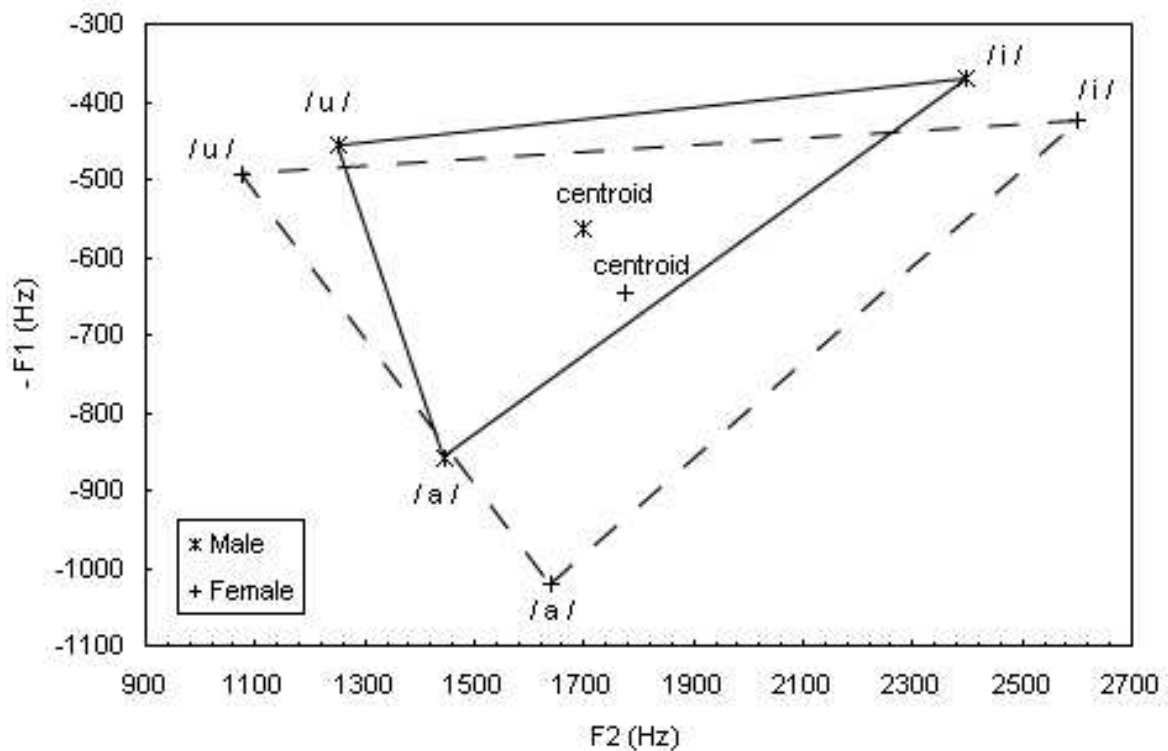


Figure 4. Acoustic vowel spaces for males and females.

Discussion

The Pearson product-moment correlation coefficients between the acoustic and articulatory ED (-0.255 for males and -0.484 for females) appear to be not strong enough to suggest acoustic vowel space as a measurement of articulation. Yet, such imperfect correlation was supportive of a non-linear articulatory-acoustic relationship. The acoustic and articulatory distances (EDs) between vowels do not seem to be proportional. This is also obviously when comparing the acoustic and articulatory vowel spaces. The articulatory vowel space exhibited a relatively shorter distance between /i/ and /a/ than between other vowels,

suggesting /i/-/a/ is an unstable region, and a small articulatory movement results in large acoustic change. In contrast, the /u/-/a/ region appears to be more stable, with large articulatory movement leading to only a small acoustic change. This phenomenon is consistent with the quantal theory advocated by Stevens (1989), who suggested the existence of stable and unstable regions in vowel production. The stable region allows imprecise articulation to achieve invariant acoustic output, defining the vowel identity, while the unstable regions define the boundaries among vowels (Stevens, 1989).

Convergent evidence may be found in the perceptually based vowel quadrilateral containing the entire Cantonese vowel inventory. Compared to the /u/-/a/ region, the /i/-/a/ region is generally more unstable. It follows that, between /i/ and /a/, the proportion of unstable region to stable region should be higher. In other words, there were more vowel boundaries and finer vowel discrimination. More vowels are expected to be found in the /i/-/a/ region than /u/-/a/ region. Referring to the vowel quadrilateral, there are 5 vowels (/y, ɪ, ε, œ, ɐ/) distributing between /i/ and /a/, and only 3 vowels (/ə, ʊ, ɔ/) between /a/ and /u/ (International Phonetic Association, 1999). In view of this, the quantal theory appears to be applicable to the perceptually based vowel quadrilateral.

The articulatory mechanism contributing to the non-linearity between acoustics and articulation was also of interest. The stable region was suggested to be formed by acoustic coupling. As the vowel constriction divided the oral cavity into two cavities, the front and

back cavities resonate respectively, and give rise to two formants of different frequencies.

However, the incomplete constriction allows the interaction between the two formants, so a change of one formant may be neutralized by a change of another. Stable region might be formed in this way, with the formant frequencies invariant to articulation changes. In between the stable regions were then the unstable regions.

The design of the present study may also be one of the factors contributing to the lower correlation between acoustics and articulation. Currently articulation data was solely represented by lingual positioning. Apart from different lingual activities, the three corner vowels were produced also with different lip gestures, with spread lips for /i/, relaxed lips for /a/, and protruded lips for /u/. These variations in lip articulation were not accounted for in the EMA kinematic measurements, although they contributed to and altered the configuration of the front resonant cavity and the resonant frequencies. In the current calculation, even the acoustic outcomes from lip articulation were assumingly attributed to the lingual articulations. This would input acoustic data irrelevant to the articulation data and as an artifact lower the correlation between acoustics and articulation.

In particular, the lip protrusion in /u/ production was suspected to have larger influence on acoustics, lengthening the front cavity in acoustic resonance and lowering the corresponding resonant frequency. Whether the front cavity affiliated with F1 or F2 remained uncertain; however, in either one of the scenarios /u/ would become more distant from /a/ in

acoustics. According to the acoustic vowel space shown in Figure 4, decreasing the F1 or F2 values of /u/ would lengthen the /u/-/a/ acoustic distance. Putting this unaccounted effect into consideration, the aforementioned findings of shorter /u/-/a/ in acoustics than in articulation will still be valid, if not more vigorous. If lip protrusion was measured and its effect of lengthening /u/-/a/ acoustic distance was taken out, the /u/-/a/ acoustic distance was expected to be even shorter than that currently observed.

Another observation worthy of discussion was the gender difference in correlation between acoustics and articulations, with males having weaker correlation which may be attributed to discrepancies in resonant properties between front and back resonant cavities. For males, their back resonant cavities were significantly lengthened upon the lowering of larynx during puberty. The front cavity and back cavity were hence expected to have different resonant properties. In the longer back cavity, a unit of articulatory movement was proportionally downscaled and led to a smaller change in acoustics. The present results showed consistent evidences. As seen in Figure 4, the /u/-/a/ acoustic distance was shorter than that of /i/-/a/, more significantly so in males. Given that /u/-/a/ region corresponded to the back cavity, the results agree that the back cavity resonance was less sensitive to articulatory changes. In males the discrepancy between the front and back cavities was larger, resulting in a weaker correlation in the articulatory-acoustic relationship.

More gender differences were present to bring further insights into the

articulatory-acoustic relationship. Referring to Figure 3, females showed shorter articulatory vowel space in the horizontal dimension (tongue advancement). This was reasonable considering the shorter vocal tract in female, and consequently the resonant cavities in vowel production were also shorter, resulting in higher resonant frequencies including F1 and F2 (Diehl, Lindblom, Hoemeke & Fahey, 1996). This notion is supported by the present results.

As shown in Figure 4, the female acoustic vowel space is apparently larger than the male acoustic vowel space, at least in the high frequency region. However, things are different in the low frequency region. The male acoustic vowel space appears to be slightly larger than the female one in the low frequency region, in particular, females exhibited a lower F2 for the vowel /u/ (see Figure 4).

It was unexpected that females produced /u/ with a lower F2 than that of males, in contradiction with the shorter vocal tract and the generally higher formant frequencies in females. Not likely to be accountable by anatomical differences, the phenomenon may be explained by the gender difference in articulatory behaviors. Female's voice is associated with higher fundamental frequency (faster average vocal fold vibration during phonation). Given that harmonics are multiples of fundamental frequency, female vowel formants contain more loosely packed harmonics (a lower resolution). To compensate for the poorer clarity in acoustic output, females may adopt a distinct set of articulatory gestures to create a more dispersed acoustic vowel space, so as to magnify the inter-vowel contrasts (Ryalls &

Lieberman, 1982). This may explain the lower F2 in /u/ produced by females, as females were using a gender-specific articulatory gesture to lower the formant frequencies. Females behaviorally attempted to lower the formants due to the fact that the anatomical configuration had already extended the acoustic vowel space at the high frequency end. Consistent with this notion the female's acoustic vowel space was not exceeded much by the male one at the low frequency end (see Figure 4).

However, one may wonder how female speakers manipulate the vocal mechanism to achieve lower formant frequencies. A possible way of reducing formants in females may be to lengthen the resonant cavity, such as /u/ with low F1 and F2. However, this was contradictory to the horizontally shorter articulatory vowel space in female (see Figure 3). It should be noted that the diagram only accounts for lingual articulation. It is speculated that lengthening of resonating vocal tract is mainly done by protruding the upper and lower lips, and/or alternatively lowering of larynx. Apparently, more information especially imaging data is needed to confirm this speculation.

Apart from altering vocal tract length, formant frequencies can be reduced by other methods. According to the perturbation theory, two approaches can be used to lower formant frequencies. One is to constrict the point of maximum velocity (the waveform node), so that the air molecules are impeded to move, and hence moving with a lower frequency (Johnson, 2003). For instance, in /u/ production the points of maximum velocity were at the lip and in

the velar region (Diehl, Lindblom, Hoemeke & Fahey, 1996). Females may have a narrower constriction at the lip and the velar region, which may yield lower resonant frequencies.

In addition to the location of constriction, the perturbation theory also suggests that dilation of point of maximum pressure (the waveform antinode) may also affect resonances. At the point of minimum velocity at waveform antinode, air molecules would have a lower tendency to move, and hence also a lower frequency in movement (Johnson, 2003). Taking the vowel /u/ as an example, the points of minimum velocity were estimated to be around the palatal region and the pharyngeal region (Diehl, Lindblom, Hoemeke & Fahey, 1996). To lower the frequency, female speakers may dilate the palatal and pharyngeal region more than male speakers.

In conclusion, results from the present study support the non-linear acoustic-articulatory relationship and suggest that lip and laryngeal gestures are also determining factors of the acoustic outcomes. Furthermore, data suggest that apart from gross movement of tongue height and frontness, fine alternations of oral cavity may be combined to significantly alter vocal tract resonance characteristics, and thus formant frequencies.

Future studies may explore the predictability and regularities in the non-linear articulatory-acoustic relationship, so that despite the current challenges, acoustics can be further exploited as a possible indicator of articulatory capabilities for clinical application. Towards this goal a possible approach was to divide the resonant cavity into various regions,

and for each region define a specific articulatory-acoustic function, in order to overcome the non-linear resonance nature along the vocal tract.

References

- Diehl, R. L. (2008). Acoustic and auditory phonetics: the adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 965-978.
- Diehl, R. L., Lindblom, B., Hoemeke, K. A., & Fahey, R. P. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, 24, 187 – 208.
- Eir Cortes, E., & Lindblom, B. (2008). *From articulatory to acoustic parameters non-stop*. Paper presented at FONETIK, Gothenburg, Sweden, June, 2008.
- Gloria, J. B., Harris, K. S., & Raphael, L. J. (1994). *Speech Science Primer – Physiology, Acoustics, and Perception of Speech* (3rd ed.). Baltimore, MD: Lippincott Williams & Wilkins.
- Goozée, J. V., LaPointe, L. L., & Murdoch, B. E. (2003). Effects of speaking rate on EMA-derived lingual kinematics: A preliminary investigation. *Clinical Linguistics and Phonetics*, 17, 375-381.
- Hashi, M., Kent, R. D., Westbury, J. R., & Lindstrom, M. J. (1996). *Relating formants and articulation in intelligibility test words*. Paper presented at the ICSLP Conference, Philadelphia, October, 1996.

Hoole, P., & Hu, F. (2004). *Tone-vowel interaction in standard Chinese*. Paper presented at International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages, Beijing, March, 2004.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Johnson, K. (2003). *Acoustic and auditory phonetics* (2nd ed.). Malden, MA: Blackwell Publishing.

Kaburagi, T., & Honda, M. (2002). Electromagnetic articulograph based on a nonparametric representation of the magnetic field. *Journal of the Acoustical Society of America*, 111, 1414-1421.

Kent, J. F., Kent, R. D., Rosenbek, J. C., Weismer, J. C., Martin, G., Sufit, R., & Brooks, B. R. (1992). Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research*, 35, 723-733.

Lee, W. S. (2005). The articulatory and acoustical characteristics of the ‘apical vowels’ in Beijing Mandarin. *Journal of the Acoustical Society of America*, 118, 2027.

Ling, F. (2007). *The articulatory and acoustic study of fricative vowels in Suzhou Chinese*. Paper presented at the ICPhS Conference, Saarbrücken, Germany, August, 2007.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T.

T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *The Journal of the Acoustical Society of America*, 92, 3078-3096.

Ryalls, J. H., & Lieberman, P. (1982). Fundamental frequency and vowel perception. *Journal of the Acoustical Society of America*, 72, 1631-1634.

Simpson, A. P. (2002). Gender-specific articulatory-acoustic relations in vowel sequences. *Journal of Phonetics*, 30, 417-435.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3 – 45.